3 The Monte Carlo Method, an Introduction

Detlev Reiter

Institut für Energieforschung - Plasmaphysik, Forschungszentrum Jülich GmbH, 52425 Jülich, Germany

This chapter presents the basic principles of stochastic algorithms, usually called Monte Carlo methods. After some historical notes, the generation of random numbers is discussed. Then, as a first non-trivial example, the concept is applied to the evaluation of integrals. More involved problems will be discussed in the two subsequent chapters of this part.

3.1 What is a Monte Carlo Calculation?

In an early lecture note (around 1960, but see also [1]) one of the pioneers of the Monte Carlo technique M.H. Kalos, quotes the two "definitions": (i) A last resort when doing numerical integration, and (ii) a way of wastefully using computer time. Today common assumptions characterize it as a numerical method involving random numbers in a significant way.

In a certain sense any large computer calculation has random aspects, due to roundoff errors. Also deterministic molecular dynamics (MD) calculations, in which the interaction of a large number of moving particles is followed by integrating Newton's equation can have random results, due to randomly chosen initial conditions and/or due to the large number of particles. These are usually excluded from that definition of Monte Carlo techniques. The involvement of randomness in Monte Carlo methods is rendered more precisely to mean deliberate use of random numbers in a calculation which has the structure of a stochastic process.

Two major areas of application are in statistical mechanics (many particle systems) and in linear kinetic (particle) transport theory. The first type of calculations are embodied in a very specific sampling technique, and are not discussed here. The second, for example traffic flow, finance, genetics, but in particular neutronics, radiation transport, cosmic rays, neutral and charged particle transport in plasmas, etc., rely on a study of many interesting stochastic processes by imitating the random processes directly on the computer. Although intuition, and the resulting high transparency of the procedure, is an important ingredient in this type of stochastic analysis, a sound mathematical basis also exists. This allows rigorous mathematical proofs to be given that certain methods actually provide solutions to certain generic mathematical equations, e.g., to Fredholm Integral Equations of second kind in case of transport problems. Distinct from most numerical schemes, in these stochastic methods error estimates are provided by the method itself rather than requiring

D. Reiter: *The Monte Carlo Method, an Introduction*, Lect. Notes Phys. **739**, 63–78 (2008)
DOI 10.1007/978-3-540-74686-7_3
© Springer-Verlag Berlin Heidelberg 2008

additional considerations. Some mathematical background and basic statistical results are also needed to analyze results of Monte Carlo simulations, for estimation of errors and for obtaining more economical approaches beyond simple simulation of nature.

There exists a vast amount of introductory literature, from a basic text-book level up to monographs focussing on very specialized applications (see, for example, [1, 2, 3, 4]), probably now hundreds of web-based lecture notes and uncounted journal articles. This present introductory chapter certainly duplicates most, if not all of that material. Our aims are to introduce the terminology, and to convey the message that Monte Carlo Methods do have a solid basis in measure theory (with the theory of probability as special case thereof). Strict mathematical proofs of convergence of the method to the exact solution exist, but also, and distinct from most numerical concepts, implementation can be be strongly guided by intuition and retain a high transparency even in very complex situations.

We will, after some short historical remarks below, start with introducing the concepts of random events, of error estimates and unbiased procedures for estimation. Practical implementations of Monte Carlo techniques rely on our ability to draw random number from any probability law we wish. Only a few, most basic facts and concepts in this regard will be repeated here in Sect. 3.2. In Sect. 3.3 the central limit theorem and variance reduction techniques will be demonstrated *at work* using the generic example for Monte Carlo methods: Integration by stochastic sampling. The relation of this very general but intuitively clear and transparent application to the mathematically and statistically more involved transport problems (Monte Carlo particle simulation) will be frequently used as guidance here and will be discussed in more detail in Chap. 5.

3.1.1 Historical Notes

Monte Carlo concepts fall into the branch of experimental mathematics. In ordinary mathematics conclusions are deduced from postulates (Deduction). In experimental mathematics conclusions are inferred from observations (Induction). Monte Carlo methods comprise that branch of experimental mathematics, which is concerned with experiments on random events (mainly random numbers). Monte Carlo methods can be of probabilistic or deterministic type.

Usually the first reference to the Monte Carlo Method is the famous needle experiment of Compte de Buffon (1733), a French biologist (1707–1788), Fig. 3.1. Buffon pointed out that if a needle of length L is tossed on a plane with parallel lines a distance D apart ($D > L$), it has probability $p = 2L/(\pi D)$ to fall such that is crosses one of the lines. Later, also Laplace suggested this procedure to determine π by counting the number of crosses n in N repetitions of the experiment. Then

$$\frac{n}{N} = \frac{2L}{\pi D} \Rightarrow \pi \approx \frac{2L}{D} \cdot \frac{n}{N} \,. \tag{3.1}$$

This historical use of Monte Carlo has all key features of the method:

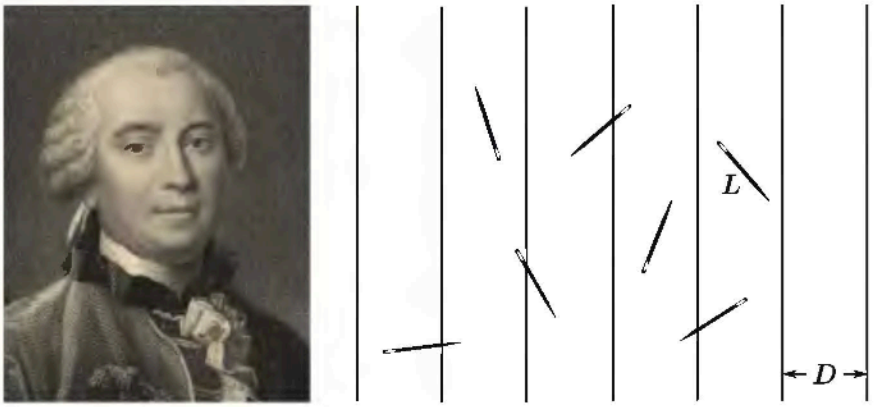


Fig. 3.1. Buffon's needles: What is the probability p, that a needle (length L), which falls randomly on a sheet, crosses one of the lines (distance D)? (**Left**: ©Copyright 1998–2003: The Regents of the University of California)

- *Convergence:* About $N = 100\,000$ trials are needed for only two digits after the comma. Convergence is slow, but foolproof.
- *Transparency:* The method is intuitively understandable, even without any mathematical reasoning.
- *Error estimates, optimization:* Error estimates and optimal choice of L, D are provided by theory of probability. (Binomial distribution, statistical variance as 2^{nd} central moment etc.).

Modern use Monte Carlo techniques, in the age of digital computers, was initiated by the pioneering work of John von Neumann and Stanislaw Ulam in thermonuclear weapon development. They are also credited for having coined the phrase *Monte Carlo*.

Many monographs on Monte Carlo Methods start with an introduction to measure theory and in particular to elementary probability theory. Although we will introduce and use the proper mathematical vocabulary too, we will, with respect to purely mathematical aspects, refer to those and largely rely upon the intuitive meaning. We refer in particular to the classic monograph by Hammersley and Handscomb [2]. This book provides a short and very readable overview of Monte Carlo. Remarkably, the theoretical foundations today remain rather similar to those from 1964, when this book was first published. Just the applications are far more sophisticated today. The illustrative examples on Monte Carlo integration and some of the advanced techniques in this present introduction will be based upon this text[1].

[1] A pdf-file of that book, which is out of print since long, can be downloaded from the internet, e.g., http://www.eirene.de/html/textbooks.html.

3.1.2 The Basic Principle

The principle is to find (estimate) mean values, i.e. expectation values, I of some system components. If a deterministic problem is to be solved, one first has to invent a stochastic system such that a mean value (= expectation value) coincides with the desired solution I of the deterministic problem.

In any case: I is a single numerical quantity of interest (not an entire functional dependence), and one might always think of I as some definite integral.

The simple intuitive interpretations are given below, but in abstract mathematical terms this stochastic model is given by the probability space (Ω, σ, p, X). Ω is a set of elementary (random) events ω, the σ-field is a set of subsets of Ω to which the measurable function p assigns a value (the probability) from the interval [0, 1], such that the Kolmogoroff axioms for a probability are fulfilled. X is a random variable on Ω, assigning a (usually real) number (or vector) to each random event, e.g.: $X(\omega) \rightarrow \mathbb{R}$, such that $I = E(X)$, the expected value of X.

The expectation value $E(X)$ and variance $\sigma^2(X)$ are defined as the first moment and second central moment, respectively, and, unless otherwise stated, we assume that they both exist

$$\begin{aligned} E(X) &:= \int_{\Omega} \mathrm{d}p \; X \; , \\ \sigma^2(X) &:= \int_{\Omega} \mathrm{d}p \; (X - E(X))^2 \; . \end{aligned} \tag{3.2}$$

Note that $E(X) = E_p(X)$, $\sigma^2(X) = \sigma_p^2(X)$, i.e., the moments of X of course depend upon the probability measure p.

A stochastic approximation to I is then obtained by producing an independent sequence of random events $\omega_i, i = 1, \ldots N$ according to probability law p and evaluating

$$E(X_N) = I_N = \frac{1}{N} \sum_{i=1}^{N} X(\omega_i) \; . \tag{3.3}$$

The estimator I_N is just the arithmetic mean of many (N) outcomes of the random experiment.

Even without any of this abstract mathematical background it is intuitively clear (see examples below) that I_N will converge to $E(X)$, hence to I by construction, as the number of samples N is increased. However the laws of large numbers and the central limit theorems of probability theory not only provide sound mathematical proofs that this Monte Carlo procedure is exact (unbiased) but also that it converges: $I_N \rightarrow I$ for $N \rightarrow \infty$, albeit slowly (with $1/\sqrt{N}$). In particular the central limit theorem of probability theory[2] asserts that the probability distribution of I_N, for large enough N, converges to a Gaussian distribution, with mean value $I = E(X)$ and

[2] See any textbook on Monte Carlo, or Probability Theory.

variance $\sigma^2(I_N) = \sigma^2(X)/N$. Hence the typical results from statistical error analysis under Gaussian distribution laws apply, e.g., also the resulting confidence levels. It is, therefore, common practice in Monte Carlo applications to quote results as

$$I \approx I_N \pm \sigma(I_N) \quad \text{or} \quad I \approx I_N \pm 2 \cdot \sigma(I_N) \,, \tag{3.4}$$

which have confidence levels of about 66% and 95%, respectively.

Of course, in applications the variance $\sigma^2(X)$ is usually even more difficult to compute than the mean value $E(X)$. It is therefore replaced by the empirical variance

$$\begin{aligned} s^2 &= \frac{1}{N-1} \sum_{i=1}^{N} [X(\omega_i) - E(X_N)]^2 \\ &= \frac{1}{N-1} \left(\sum_{i=1}^{N} X^2(\omega_i) - \frac{1}{N} \left[\sum_{i=1}^{N} X(\omega_i) \right]^2 \right) \end{aligned} \tag{3.5}$$

and one has also, under the assumptions made, for large sample size N

$$\begin{aligned} s^2 &\to \sigma^2 \,, \\ \sigma^2(I_N) &\approx s_N^2 = \frac{1}{N} \, s^2 \,. \end{aligned} \tag{3.6}$$

Hence, for large enough N, in the Gaussian based error estimates (3.4) σ can safely be replaced by s_N, at least for large sample size $N \gtrsim 100$. In the opposite case $N \lesssim 100$ Student's t-distribution should be employed in error analysis instead.

3.2 Random Number Generation

The Monte Carlo method rests on our ability to produce random numbers drawn from any particular probability distribution $F(x)$, or, if it exists, from the probability density function (pdf) $f(x)$, with $F(x) = \int_{-\infty}^{x} \mathrm{d}t \, f(t)$.

Examples are wetting of a surface by rain (uniform distribution), radioactive decay (Poisson distribution), or the distribution of velocities of molecules in a gas (Gaussian distribution). In general, the probability law must be known, either from theory, experiment or plausibility.

A theorem from measure theory states:

Theorem 1. *Each (probability) measure μ can be decomposed into a weighted sum $\mu = p_1 \, \mu_c + p_2 \, \mu_d + p_3 \, \mu_a$ of three parts:*

(i) Part one has a continuous distribution (with probability density $f(x)$).
(ii) Part two has a discrete distribution.
(iii) Part three is a pathological contribution.

The third part is required for the abstract mathematical case only (general measurable spaces, σ-algebras, ...), but it does not occur in practical Monte Carlo applications. This means, for any distribution law arising in an application we can obtain random numbers in two steps: First a random decision (based on the two remaining weighting factors p_1, p_2) whether the continuous or the discrete distribution is to be sampled, and second then generating a random number from the chosen distribution μ_c or μ_d. We will show below that for both cases, continuous and discrete distributions, general procedures for random number generation exist, at least in principle. We refer to the standard reference on the production of nonuniform random numbers [5]. This book deals with the myriad number of ways to transform the uniform random numbers into anything else one might want. Also the first section (pp. 1–193) of [6] is a very comprehensive introduction to random number generation.

3.2.1 Uniform Random Numbers

Uniform random numbers are the basis for generation of random numbers with all other distribution laws. A random variable is uniformly distributed on an interval [a, b], if the distribution density f is

$$f(x) = \frac{1}{b-a}\chi_{[a,b]}\ , \qquad x \in \mathbb{R} \tag{3.7}$$

with $\chi_{[a,\,b]} = 1$ if x is in the interval $[a,\ b]$ and $f(x) = 0$ elsewhere.

The classical method to generate uniform random numbers on $[0, 1]$ is by so called linear congruential random number generators, which are defined by the recursion

$$\xi_{n+1} = [a\,\xi_n\ +\ b] \mod m \tag{3.8}$$

Here a is a magic multiplicand, m if often chosen to be the largest integer representable on the machine ($m = 2^{32}$, etc.), and b should be prime to m. Proofs for particular choices of (large) parameters a and m that the generator achieves the largest possible period of $m - 1$ different random numbers are quite cumbersome. Optimal parameter choices are typically found experimentally, see again [6]. The finite periodicity limits precision only in very large calculations, e.g. on modern massively parallel computing systems. A rather subtle issue is also *independence* of an entire sequence of random numbers (loc.cit.).

3.2.2 Non-uniform Random Numbers

As stated above we only need to consider discrete distributions and continuous distributions.

Finding random numbers with a given discrete distribution is trivial: Let a discrete distribution, with k elementary outcomes labelled by natural numbers $\{0, 1, 2, \ldots, k\}$, be given by

$$\begin{aligned} P(X = i) &= p_i \geq 0 \,, \\ \sum_{i=0}^{k} p_i &= 1 \,, \\ F(i) = P(X \leq i) &= \sum_{j=0}^{i} p_i \end{aligned} \tag{3.9}$$

with $P(X = i)$ the probability of event i. F is the (cumulative) distribution. Let ξ be a uniform random number on $[0, 1]$, then the random variable X with $X = i$ if $F(i-1) < \xi \leq F(i)$ is distributed according to F.

3.2.2.1 Inversion Method

The inversion method provides random samples z from a distribution F by converting uniform random numbers ξ. This is simply done by setting

$$z := \min\{x | F(x) \geq \xi\} \sim F \,. \tag{3.10}$$

If F is strictly monotonous, then $z = F^{-1}(\xi)$. For example, if $f(x)$ is the distribution density function (pdf) to be generated, then first find the cumulative function $F(x) = \int_{-\infty}^{x} \mathrm{d}t \; f(t)$, pick a uniform random number ξ on $[0, 1]$ and set $\xi = F(z)$ and finally invert this to find random number z, which is then distributed according to $f(x)$.

The same transformation rules as for any density function apply also for a pdf. Hence the general strategy is: Try to transform a given pdf $f(x)$ to another distribution $\widetilde{f}$, such that the inverse of the new cumulative distribution $\widetilde{F}$ is explicitly known. Then apply the method of inversion and transform back. Figure 3.2 illustrates the method of inversion for the normal (Gaussian) distribution

$$\begin{aligned} \phi(x) &= \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2} \,, \\ \Phi(x) &= \int_{-\infty}^{x} \mathrm{d}t \; \phi(t) = \frac{1}{2} \left[1 + \mathrm{erf}\left(\frac{x}{\sqrt{2}} \right) \right] \,. \end{aligned} \tag{3.11}$$

Unfortunately, the Gaussian error function $\mathrm{erf}(x)$ and hence $\Phi(x)$ cannot be inverted in closed form. We will show how to generate Gaussian random numbers, even without numerical inversion, further below.

From this procedure follows directly the natural and best format for storing (also multi-dimensional) tabulated data for random sampling in Monte Carlo applications: Form the inverse cumulative distribution function $F^{-1}(x)$ (i.e.: the quantile function) and store this for x uniformly spaced in $[0, 1]$. Then take ξ from a uniform distribution on $[0, 1]$ and find $F^{-1}(\xi)$ by interpolation in this table.

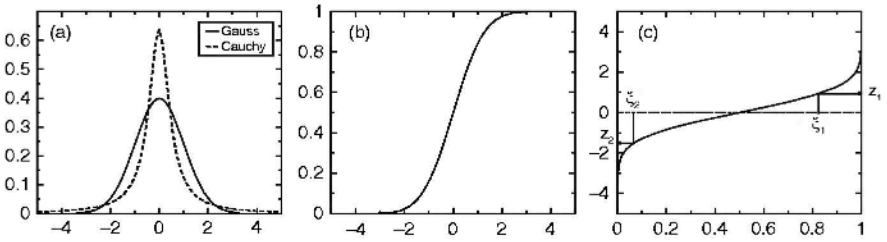


Fig. 3.2. (**a**) Comparison of Cauchy (*dashed line*) and normal distribution (*solid line*). (**b**) Cumulative distribution function $\Phi(x)$ of normal distribution (3.11), (**c**) Inverse cumulative distribution of normal distribution $\Phi^{-1}(\xi)$. Uniform random numbers ξ_1, ξ_2 (abscissa) are converted to random numbers z_1, z_2 from a normal distribution (ordinate)

3.2.2.2 Rejection

Another general method for generating non-uniform random numbers is the rejection method (J.v. Neumann, 1947). This method is *always* applicable, although it may sometimes be rather inefficient. For distributions with finite support, i.e., $f(x) \neq 0$ only on a finite domain M (say, $M = [a, b]$), find the maximum c of $f(x)$, sample a random pair (ξ_1, ξ_2) with ξ_1 uniform on M and ξ_2 uniform on $[0, c]$. If $\xi_2 \leq f(\xi_1)$, accept ξ_1. Otherwise reject this pair and pick a new pair. Repeat this procedure until a pair is accepted. Clearly, the efficiency of this method (e.g. measured as average number of accepted random pairs to number of pairs produced) may be quite poor, in particular if the distribution $f(x)$ has sharp maxima.

A more general, and sometimes more efficient rejection method, working even on infinite sampling domains M, results if one finds a second distribution $g(x)$ and a numerical constant c such that $f(x) \leq c \cdot g(x)$. Again find a pair (z_1, z_2) of random numbers, however with z_1 not sampled uniformly on M but from distribution $g(z)$ instead. z_2 is uniform on the interval $[0, c]$. The random variable z_1 is accepted if $z_2 \leq f(z_1)/g(z_1)$. Otherwise a new pair (z_1, z_2) is generated. See Chap. 5 for an important application in particle simulation.

3.2.2.3 Examples

3.2.2.3.1 Inversion

Important examples in which the inversion method can be applied are, e.g., the exponential distribution (of the mean free flight length of radiation in matter), the cosine distribution of polar emission angles against surface normals, the surfaces crossing Maxwellian flux distribution $f(v_\perp) \propto v_\perp f_{\text{Maxw}}(v_\perp)$ of normal velocity components of gas molecules with Maxwellian velocity distribution (f_{Maxw}). We explicitly illustrate the inversion method here for the Cauchy distribution: The Cauchy distribution, see Fig. 3.2(a), in physical applications also called Lorentz distribution, is an example of a distribution function that has no moments. It arises often in radiation transfer, e.g., as line-shapes of naturally- or Stark broadened lines or in other resonance phenomena

$$f_C(x) = \frac{c}{\pi}\frac{1}{(x-b)^2 + c^2} \,. \tag{3.12}$$

Here b is the median (line shift), and c is the half width at half maximum (HWHM). Generating random number with a Cauchy distribution is usually done by inversion. First transform to a standardized Cauchy, by $s = (x - b)/c$. The cumulative distribution is then given as

$$F_C(x) = \frac{1}{\pi}\int_{-\infty}^{x} \frac{1}{s^2+1}\,\mathrm{d}s = \frac{1}{2} + \frac{1}{\pi}\arctan\left(\frac{x-b}{c}\right) \,. \tag{3.13}$$

Therefore the random number $z = b + c \cdot \tan\{\pi(\xi - 1/2)\}$, with ξ a uniformly distributed random number on $[0, 1]$, has a Cauchy (b,c) distribution.

3.2.2.3.2 The Box Muller Method for Gaussian Random Numbers

Because the Gaussian error function cannot be inverted in closed form, the following combination of transformation, rejection and inversion method is typically applied: Not one, but two independent normally distributed random numbers (z_1, z_2) are produced by first transforming random variables Z_1, Z_2 from cartesian to polar coordinates R, Φ. The angle Φ is then uniform in $[0, 2\pi]$. Only $\cos(\Phi)$ and $\sin(\Phi)$ are needed, and a rejection method (comparing a unit circle and a surrounding square) can be used for them. The variable R has, due to the Jacobian of the transformation, a Gaussian flux distribution (see above) rather than a Gaussian itself, and this can be directly generated by the Method of Inversion. Transforming back $Z_1 = R \cdot \cos(\Phi)$ and $Z_2 = R \cdot \sin(\Phi)$ provides a pair of independent Gaussian random numbers.

3.3 Integration by Monte Carlo

Integration by Monte Carlo is a stochastic method for the deterministic problem of finding an integral, which in sufficiently complex high dimensional situations can be competitive or even superior to numerical methods.

Let's consider the source rate of particles (likewise, of momentum, heat, etc.) in a macroscopic system (e.g., a fluid flow), in which these particles (microscopic objects) are ruled by a kinetic, i.e. microscopic (Boltzmann) equation. Examples are chemical sources (particle, momentum, energy) in plasma chemistry, or radiative heat source in case of radiation transfer theory.

Such terms then read

$$I = \int_V \mathrm{d}x\; g(x) f(x) := \int_V \mathrm{d}f\; g(x) \,. \tag{3.14}$$

Here f is the one particle distribution (density) function $f(\boldsymbol{r}, \boldsymbol{v}, i, t)$ or $f(x)$, where the state x of the relevant phase-space may, e.g., be characterized by a position vector $\boldsymbol{r}$, a velocity vector $\boldsymbol{v}$, the time t, i.e. continuous variables, and further a discrete

chemical species index i, also for example for internal quantal states. $g(x)$ is again some weighting function determined by the particular moment of interest. In mathematical terms one would refer to this as Lebesgue-Stieltjes Integral of measurabel function $g(x)$ with respect to (probability) measure defined by distribution density $f(x)$.

We will discuss Integration by Monte Carlo using the example from [2]: Let the integration domain V be the unit interval $[0, 1]$, $f(x)$ the uniform distribution on $[0, 1]$ (i.e.: $f(x) = 1$ on $[0, 1]$, and $f(x) = 0$ elsewhere) and $g(x) = (\exp(x) - 1)/(\mathrm{e} - 1)$. Clearly,

$$I = \int_0^1 \frac{\mathrm{e}^x - 1}{\mathrm{e} - 1} \mathrm{d}x = 0.418\,0\ldots\,. \tag{3.15}$$

We will now integrate this same function by Monte Carlo. Our first method does not require any theory, but instead, inspired by Buffon's needle experiment, we will just use pairs ξ_1, ξ_2 of independent uniform random numbers and compare the known area (the unit square $[0, 1] \times [0, 1]$) with the unknown area I, which is the area underneath function $g(x)$, in [0,1]. I.e., we count a *hit* if the point defined by the pair of random numbers is under the curve $g(x)$, and a *miss* otherwise.

As can clearly be seen on Fig. 3.3 the ratio of hits to total number of samples converges to the exact values of the integral, as expected, and also the statistical error, indicated as empirical standard deviation s_N, (3.5) scales with $1/\sqrt{N}$ as expected.

Of course such a Monte Carlo integration method is patently foolish. By this method we have, in principle, replaced the single integral over function g by a double integral over the area between abscissa and function $g(x)$. The conventional text-book method (crude Monte Carlo) can be obtained from this one by the observation that once the first random number ξ_1 of the pair is known, we do not have to rely upon ξ_2 to decide about counting zero or one. Given ξ_1, then an one will be counted with probability p: $p = g(\xi_1)$. Hence instead we can use that (conditional) expected value p of the binomial distribution $b(1, p)$ directly. This is, admittedly, a quite obscure explanation for something really trivial. But it is also the underlying idea behind a powerful variance reducing Monte Carlo technique known under different names in different areas of application: Conditional expectation estimator (in neutron shielding), [4], averaging transformation (transfer theory, mainly in Russian literature), [7], or energy partitioning method in radiative heat transfer [8].

This method is opposite to randomization: We have replaced a sampled result (zero or one) by its expectation value. In our particular example we have carried out one of the two integrations analytically, conditional on the outcome ξ_1. The second random number is not needed at all in this particular trivial case but this not the relevant point. What is important also in general terms is that one (generally: some) of the two (generally: many) integrals has been done analytically, and only the remaining ones by random sampling. The general rule is: Always try to do as many integrations analytically or numerically and resort to Monte Carlo only for the rest. In particle transport theory this concept will lead to powerful hybrid methods combining information gained analytically (or numerically) and stochastically, bridging

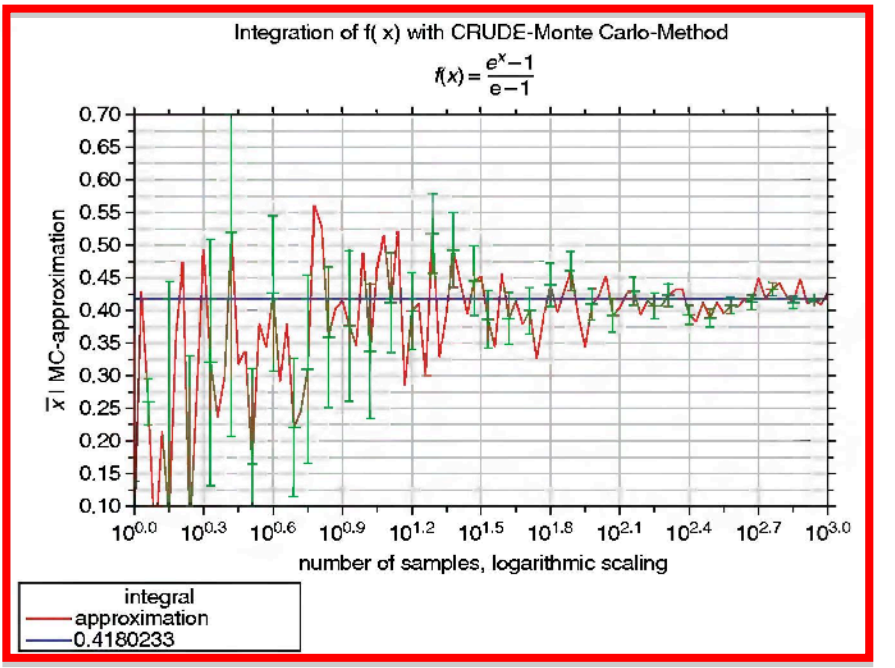


Fig. 3.4. Evaluating Integral of $(\exp(x) - 1)/(e - 1)$ on [0,1], method: crude Monte Carlo

statistical noise is reduced solely by modifying (smoothing) the estimator $g(x)$, in importance sampling the underlying random variable (or random process) $f(x)$ is altered to another one, $\widetilde{f}(x)$ in order to achieve variance reduction. A compensating weight correction factor in introduced in the estimator to maintain the same mean value $I = E(g(X))$

$$I = \int_V g(x) \cdot f(x)\mathrm{d}x = \int_V g(x)\frac{f(x)}{\widetilde{f}(x)}\widetilde{f}(x)\mathrm{d}x = \int_V \widetilde{g}(x)\widetilde{f}(x)\mathrm{d}x \,. \qquad (3.16)$$

Hence we have $\widetilde{g}(x) = g(x)f(x)/\widetilde{f}(x)$. The name of this method, importance sampling originates from the special techniques often used to find optimal biassing schemes (i.e.: $\widetilde{f}(x)$) of the random process, in particular in transfer theory. A more general, but also somewhat imprecise terminology would refer to this concept as non-analog Monte Carlo, as compared to the analog Monte Carlo scheme. In the latter the underlying probability distribution law is directly taken from the application, whereas in the former one uses a different distribution, motivated by practical, economical or other reasons, and statistical weights to compensate this.

As seen from (3.16), the value of I is independent of how the integrant is decomposed into a product of a probability density and a response function, but the variances, $\sigma_f^2(g)$ and $\sigma_f^2(\widetilde{g})$, certainly can be different.

Let's take, again, our example, to illustrate the concept: In order to reduce the variance $\sigma^2_{\tilde{f}}(\tilde{g})$ of $\tilde{g}$ with respect to probability law $\tilde{f}$ we should try to make $\tilde{g}$ as constant as possible on [0,1]. The Taylor expansion of our particular function $g(x)$ indicates that the ratio $\tilde{g}(x) = g(x)/x$ should be more constant than $g(x)$ itself. Hence we try $\tilde{f}(x) \propto x$, i.e., $\tilde{f}(x) = 2x$ so that $\tilde{f}(x)$ is normalized to one on [0,1].

Our importance sampling procedure to evaluate I now proceeds as follows: Draw random numbers $\tilde{\xi}$ from $\tilde{f}(x)$. By the method of inversion, this is done by setting $\tilde{\xi} = \sqrt{\xi}$, with ξ a uniform random number on [0,1]. Then, again, form the arithmetic average of many (N) random variables $\tilde{g}(\tilde{\xi})$. Figure 3.5 shows the result of such an integration, again vs. N. Clearly the convergence is (i) to the correct value, (ii) still only $\propto 1/\sqrt{(N)}$, but (iii) the error bars s_N are much smaller than in both previously discussed Monte Carlo integration methods.

Again, it needs to be pointed out that the efficiency of the procedure is neither determined by the variance, not by N per CPU-time, but only by the figure of merit: variance per CPU time. And hence, importance sampling, more generally, non-analog sampling, can go both ways in Monte Carlo. Its performance has to be assessed on a case by case basis.

As a general observation, one should note that in non-analog Monte Carlo schemes the error assessment simply based upon the empirical variance, and error bars obtained from the central limit theorem, can be less reliable than in analog simulations. Although the variance may be decreased by a clever importance sampling

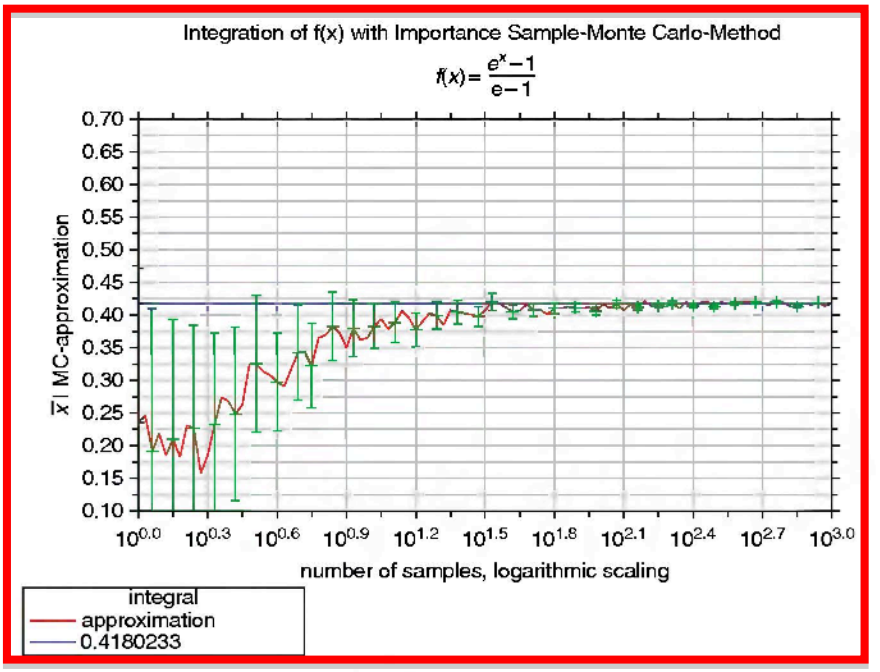


Fig. 3.5. Same integral as in Fig. 3.4, method: importance sampling Monte Carlo

method, the variance of the variance may increase, thus invalidating conventional error bar estimates, see [9].

As in the case of conditional expectation Monte Carlo we can design an extreme case of importance sampling with zero statistical error after only one sample: Let us set $\widetilde{f}(x) = g(x)/I$, hence: $\widetilde{g}(x) = I = \text{const}$. Monte Carlo integration proceeds by sampling from this distribution $\widetilde{f}(x)$ which, in case of our particular example can be done by the rejection technique. Then, independent of the sampling, I is scored. Unfortunately we needed the knowledge of the final result I already to design this perfect zero variance scheme.

3.3.0.5 δf Monte Carlo

Finally we use our simple integral to illustrate the concept of the δf Monte Carlo method, which is widely used in kinetic particle transport simulations. Starting point is the idea to split the unknown parameter into a large known *nearby* quantity and small unknown perturbation. In particle simulations this can also be the single particle distribution function $f(x)$ solving some kinetic equation or moments of this pdf. In near equilibrium situations we have

$$f(x) = f_{\text{equil}}(x) + \delta f(x) \tag{3.17}$$

with, for example, the Maxwellian equilibrium distribution f_{equil} and a small perturbation δf. It can then be advantageous to solve, by Monte Carlo sampling, only for δf rather than for the full distribution.

So let us consider our integral again, and write, accordingly, $I = I_0 + \delta I$ with I_0 the known part

$$I_0 = \frac{1}{\mathrm{e}-1}\int_0^1 \mathrm{d}x \;\; g_0(x) = \frac{1}{\mathrm{e}-1}\int_0^1 \mathrm{d}x \;\left(x + \frac{1}{2}x^2\right) = \frac{2}{3}\frac{1}{\mathrm{e}-1} \tag{3.18}$$

and δI the rest. Clearly,

$$\delta I = \int_0^1 \mathrm{d}x \; \frac{\mathrm{e}^x - 1 - x - x^2/2}{\mathrm{e}-1} \; . \tag{3.19}$$

Figure 3.6 shows the result of the estimate for I, with I_0 known and δI evaluated by crude Monte Carlo. Clearly by eliminating a large, known, contribution to I the relative errors of the estimates for any given sample size N are greatly reduced as compared to previous methods.

This method is also related to the so called correlation sampling technique, in which one would evaluate both I and I_0 by Monte Carlo techniques, but using the same random numbers. Both estimates are then positively correlated and the statistical precision of the Monte Carlo estimate for the difference δI can be substantially better than in independent estimates of I and I_0 or of I alone.

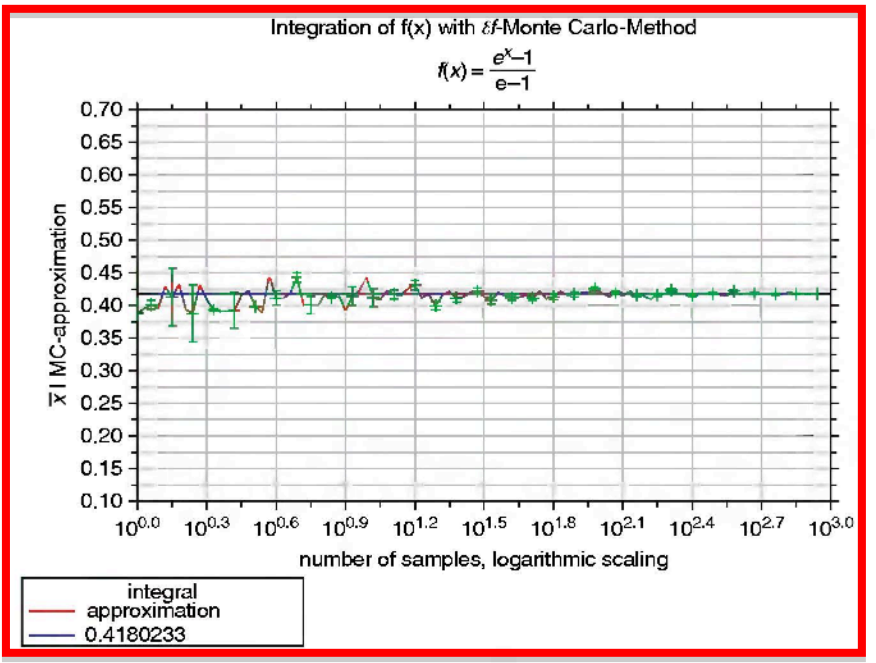


Fig. 3.6. Same integral as in Fig. 3.4, method: δf Monte Carlo

3.4 Summary

The purpose of this introduction was to show that random numbers can be generated from any given probability density distribution, and that Monte Carlo Methods can be regarded as stochastic (rather than numerical) procedures for integration. Monte Carlo consists of inventing a random game such that the expected value of a proper random variable is exactly equal to the parameter which is to be computed. Averaging over repeated independent Monte Carlo samples from that game converges (in the proper measure theoretical sense) to the desired solution.

The additional complication arising in many particle physics applications and in transfer theory is due to one fact only: Distinct from the material in this present chapter the sampling distribution $f(x)$ is sometimes not known explicitly. Instead it will be given only implicitly as solution of a, usually, very complicate equation (e.g.: the Boltzmann equation, the Fokker-Planck equation, etc.). We will see that this extra complication can be dealt with by sampling from certain stochastic processes (generating particle trajectories), rather than sampling from a given pdf, see Chap. 5. But the rest: Estimation of multi-dimensional integrals, the unbiased nature of the method, proof of convergence, error bars, variance reduction methods, remain essentially the same as in this present introduction.

References

1. H. Kalos, P.A. Whitlock, *Monte Carlo Methods, Vol. I: Basics* (Wiley-Interscience Publications, John Wiley and Sons, New York, 1986) 63, 64
2. J.M. Hammersley, D.C. Handscomb, *Monte Carlo Methods* (Chapman and Hall, London & New York, 1964) 64, 65, 72
3. R.Y. Rubenstein, in *Wiley Series in Probability and Mathematical Statistics* (John Wiley and Sons, New York, 1981) 64
4. J. Spanier, E. Gelbard, *Monte Carlo Principles and Neutron Transport Problems* (Addison Wesley Publication Company, 1969) 64, 72
5. L. Devroye, *Non-Uniform Random Variate Generation* (Springer-Verlag, Berlin Heidelberg New York, 1986) 68
6. D.E. Knuth, in *Seminumerical Algorithms*, Vol. 2 (Addison Wesley, Reading, 1998) 68
7. G. Mikhailov, *Optimization of Weighted Monte Carlo Methods* (Springer Verlag, Berlin Heidelberg New York, 1992) 72
8. A. Wang, M.F. Modest, J. Quant. Spectrosc. R. A **104**, 288 (2007) 72
9. K. Noack, Ann. nucl. Energy **18**(6), 309 (1991) 76

Kapitola 4

Základy metody Monte Carlo

4.1 Monte Carlo integrace

Jak již bylo řečeno, smyslem počítačových simulací je generovat konfigurace systému mnoha částic a tyto konfigurace pak použít ke stanovení různých termodynamických či strukturních veličin, což obvykle představuje výpočet nějaké střední hodnoty. Metoda MC generuje konfigurace právě s ohledem na efektivní výpočet středních hodnot. Název metody pak pochází z toho, že na rozdíl od deterministické MD používá generátor náhodných čísel.

■ Generátor náhodných čísel je fyzikální zařízení nebo počítačový kód, který produkuje náhodná čísla s danými statistickými vlastnostmi. Podrobnější informace najde čtenář v dodatku 12.1.

Jako úvodní příklad metody MC si ukažme výpočet čísla π, viz algoritmus 4.1. Budeme střílet náhodně body do jednotkového čtverce a stanovíme, jestli bod padl do čtvrtkruhu se středem v jednom vrcholu čtverce. Za předpokladu rovnoměrného rozložení platí

$$\frac{\text{počet bodů ve čtvrtkruhu}}{\text{počet všech bodů}} \approx \frac{\pi}{4}. \tag{4.1}$$

Při počítačové realizaci vygenerujeme dvě náhodná čísla, která budou reprezentovat bod ve čtverci; je-li generátor kvalitní, neměla by tato čísla být korelovaná. Je-li vzdálenost tohoto bodu od počátku souřadnic menší než jedna, započteme jedničku do počtu úspěšných pokusů. Tento elementární pokus musíme opakovat mnohokrát, abychom dostali co nejpřesnější výsledek; přitom platí, že chyba je nepřímo úměrná druhé odmocnině počtu pokusů.

Princip metody Monte Carlo ve statistické fyzice vysvětlíme na simulaci v kanonickém NVT souboru. Simulace v jiných souborech budou uvedeny v samostatné kapitole.

Uvažujme výpočet střední hodnoty funkce X v kanonickém souboru, rov. (2.10), a přepišme integrál do diskrétních proměnných (např. pomocí lichoběžníkového pravidla), abychom ho mohli vyčíslit na počítači:

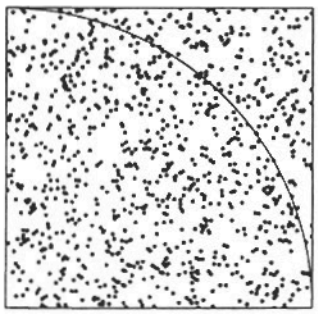
$$\langle X \rangle = \frac{\sum_{k=1}^{n} X(\mathbf{r}^{N,(k)}) \exp[-\beta U(\mathbf{r}^{N,(k)})]}{\sum_{k=1}^{n} \exp[-\beta U(\mathbf{r}^{N,(k)})]}, \tag{4.2}$$

```
INTEGER n celkový počet bodů
INTEGER i
INTEGER nu počet bodů ve čtvrtkruhu
REAL x,y souřadnice bodu ve čtverci
REAL rnd() funkce vracející náhodné číslo v intervalu (0,1)

nu := 0
FOR i := 1 TO n DO
   x := rnd()
   y := rnd()
   IF x*x+y*y < 1 THEN nu := nu + 1

PRINT "pi=", 4*nu/n

Chybu (přesně: odhad směrodatné odchylky) vypočteme z δf = ⟨(f − ⟨f⟩)²/(n − 1)⟩, kde f = 1
s pravděpodobností π/4 ≐ nu/n a f = 0 s pravděpodobností 1 − π/4 ≐ 1 − nu/n.
PRINT "chyba=", 4*sqrt((1-nu/n)*(nu/n)/(n-1))
```



Algoritmus 4.1: Výpočet čísla π metodou Monte Carlo.

kde $\mathbf{r}^{N,(k)}$ značí symbolicky k-tý bod v diskrétním souboru bodů. I kdybychom rozdělili integrační interval každé proměnné pouze na 10 dílků, dostaneme již při $N = 100$ nesmyslně velký počet konfigurací, $n = 10^{300}$. V praxi tedy nemůžeme uvažovat při výpočtu $\langle X \rangle$ všechny konfigurace, ale pouze jejich nějakou vybranou podmnožinu. Otázkou pak je, jak nejefektivněji tuto podmnožinu vybrat.

Nejběžnější metodou je vybírat tyto konfigurace zcela náhodně, což je případ známé matematické metody Monte Carlo výpočtu mnohonásobných integrálů (neboli tzv. naivní metody Monte Carlo) [32].

■ Pro odhad integrálu funkce $f(x_1, \ldots, x_D)$ přes oblast Ω v D-rozměrném prostoru platí:

$$\int_\Omega f(x_1, \ldots, x_D)\, \mathrm{d}x_1 \ldots \mathrm{d}x_D \approx \frac{|\Omega|}{n} \sum_{k=1}^{n} f(x_1^{(k)}, \ldots, x_D^{(k)}), \tag{4.3}$$

kde $(x_1^{(k)}, \ldots, x_D^{(k)})$ značí k-tý náhodný bod v oblasti Ω, jejíž D-objem je $|\Omega|$. V úvodním příkladu výpočtu čísla π jsme integrovali funkci $f(x, y) = \{1 : x^2 + y^2 < 1,\ 0 : x^2 + y^2 > 1\}$ přes $\Omega = [0, 1] \times [0, 1]$.

V případě výpočtu střední hodnoty systému mnoha částic, rov. (4.2), však tato metoda zcela selhává z jednoho prostého důvodu: náhodný výběr nedělá rozdíl mezi konfiguracemi, které mají velkou pravděpodobnost výskytu a tudíž podstatně přispívají k hodnotě $\langle X \rangle$, a málo pravděpodobnými či přímo nemožnými. Nejjednodušeji se tato skutečnost demonstruje na případu systému N tuhých koulí. Generovat náhodně konfigurace tohoto systému znamená náhodně volit polohy N koulí v objemu V (náhodně vybírat konfiguraci $\mathbf{r}^{N,(k)}$ v (4.2)) a vypočítat vždy součin funkce X s Boltzmannovým faktorem $\exp(-\beta U)$. Pro libovolnou konfiguraci však s velikou pravděpodobností (pro vysoké hustoty téměř s jistotou) dojde k tomu, že alespoň dvě koule se budou protínat, a tedy Boltzmannův faktor bude nula. Pravděpodobnost získání možné konfigurace je téměř nula a výraz (4.2) nebude vůbec definován (dělení nuly nulou).

Řešení, které se nabízí k odstranění tohoto problému, je toto: při výpočtu střední hodnoty nebudeme uvažovat libovolné konfigurace, ale přednostně ty, které podstatně přispívají k hodnotě integrálu (anglicky *importance sampling*). Otázkou pak je

1. jak toto realizovat a
2. dostaneme-li správný odhad $\langle X \rangle$.

Pokud se týká problému (1), je samozřejmě obtížné vytvořit možnou (dostatečně pravděpodobnou) konfiguraci pouhým vkládáním částic do prázdného prostoru (viz výše uvedený příklad). Máme-li však nějakou (dostatečně pravděpodobnou) konfiguraci, pak by již neměl být takový problém z této konfigurace vytvořit jinou dostatečně pravděpodobnou konfiguraci. Intuitivně lze tedy navrhnout následující schéma, které pro jednoduchost vysvětlíme opět na systému tuhých koulí. V tomto případě je totiž Boltzmannův faktor buď nula nebo jedna a tedy každá konfigurace, ve které nedochází k překryvu žádných koulí, má stejnou pravděpodobnost výskytu, zatímco konfigurace s překryvem se nikdy nevyskytnou. Předpokládejme tedy, že se nám podařilo nějakým způsobem vytvořit možnou konfiguraci. Změníme-li nyní polohu jedné (nebo i více) koulí tak, že opět nedojde k překryvu, dostaneme další možnou konfiguraci, a tak můžeme pokračovat a vytvořit posloupnost konfigurací, která bude naší vybranou posloupností v kanonickém souboru.

Jak uvidíme dále, lze výše uvedený princip rozšířit a najít vhodnou vybranou posloupnost konfigurací k výpočtu střední hodnoty i ve složitějším případě, kdy Boltzmannův faktor nabývá více hodnot než jen dvou. Rigorózně se tento problém řeší pomocí Markovových řetězců, což nakonec zajistí i správnost výsledku (bod 2).

4.2 Boltzmannovské vzorkování

4.2.1 Markovovy řetězce

Markovův řetězec je posloupnost náhodných veličin (událostí, jevů) $\mathcal{S}^{(k)}$, $k = 1, \ldots, \infty$, které se vybírají z jisté (pro jednoduchost konečné) množiny stavů (pro nás: konfigurací) $\{A_i\}$, $i = 1, \ldots, M$. Výskyt stavů není přitom nezávislý, ale událost, kterou pozorujeme v „čase“ $k+1$, závisí na tom, jaká událost byla pozorována v čase k. Konkrétně, jestliže se v čase k vyskytne událost A_i s pravděpodobností $\pi_i^{(k)}$ (tj. náhodná veličina $\mathcal{S}^{(k)}$ nabývá hodnoty A_i s pravděpodobností $\pi_i^{(k)}$), pak v čase $k+1$ se událost A_j vyskytne s pravděpodobností, pro kterou platí

$$\pi_j^{(k+1)} = \sum_{i=1}^{M} \pi_i^{(k)} W_{i\to j}\,, \tag{4.4}$$

nebo zapsáno vektorově

$$\boldsymbol{\pi}^{(k+1)} = \boldsymbol{\pi}^{(k)} \cdot \mathbf{W}\,, \tag{4.5}$$

kde matice $\mathbf{W}$ je tzv. *matice přechodu*, jejíž prvky $W_{i\to j} \equiv W(A_i \to A_j)$ mají fyzikální význam pravděpodobnosti přechodu ze stavu A_i do stavu A_j (jsou tedy nezáporné) a jež musí splňovat normovací podmínku

$$\sum_{j=1}^{M} W_{i\to j} = 1 \quad \text{pro všechna } i\,. \tag{4.6}$$

Tato podmínka znamená, že z konfigurace A_i vznikne (s pravděpodobností jedna) jedna z konfigurací A_j, $j = 1, \ldots M$.

■ Na rozdíl od Markovova *řetězce*, kde index posloupnosti k nabývá pouze diskrétních hodnot, můžeme definovat i Markovův *proces*, ve kterém máme spojitý čas. Místo $\pi_i^{(k)}$ napíšeme tedy $\pi_i(t)$. Stále však předpokládáme, že pravděpodobnosti výskytu konfigurací v čase t závisí pouze na „infinitezimálně předcházející" konfiguraci v čase $t - \mathrm{d}t$. Rov. (4.6) přepíšeme do tvaru $\pi_j(t + \mathrm{d}t) - \pi_j(t) = \sum_{i=1}^{M} \pi_i(t) W_{i\to j} - \pi_j(t) \times 1$, kde za 1 dosadíme z (4.6). Po vydělení $\mathrm{d}t$ dostaneme tzv. *řídící rovnici* (master equation)

$$\frac{\mathrm{d}\pi_j(t)}{\mathrm{d}t} = \sum_{i=1}^{M} \pi_i(t) W_{i\to j} - \pi_j(t) \sum_{i=1}^{M} W_{j\to i}\,, \tag{4.7}$$

kde elementy $W_{i\to j}$ mají nyní význam rychlosti změny stavu a neplatí pro ně (4.6) – normalizace je zajištěna druhým členem v (4.7).

Markovův proces i řetězec jsou speciálním případem obecného *stochastického* neboli *náhodného procesu*, což je systém náhodných veličin definovaných na jistém pravděpodobnostním prostoru s hodnotami v jistém měřitelném prostoru. Je třeba rozlišovat mezi *procesem*, což je systém náhodných veličin, a jeho *realizací*, nazývanou též trajektorie. Např. pro výše definovaný Markovův řetězec je jedna konkrétní posloupnost konfigurací, např. $\{A_7, A_{13}, A_7, A_{13}, \ldots\}$, jednou realizací, která se vyskytne s jistou pravděpodobností[1] v řetězci $\{\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \mathcal{S}^{(3)}, \mathcal{S}^{(4)}, \ldots\}$. Teprve když uvažujeme pravděpodobnosti *všech* realizací, můžeme mluvit o procesu!

Máme-li Markovův řetězec, můžeme si položit řadu otázek. Například, vyjdu-li ze stavu A_i, jaká je pravděpodobnost toho, že po k krocích dojdu do stavu A_j? Jak bude záviset výskyt stavu A_j na k? Tyto otázky osvětlíme nejlépe na příkladu.

Mám v kanceláři počítač. Ten mi vždy funguje, ale horší je to se sítí, ta někdy funguje, někdy ne. Dlouhodobým pozorováním jsem zjistil, že

1. funguje-li síť dnes, je 60% pravděpodobnost, že bude fungovat i zítra;
2. nefunguje-li síť dnes, pak s pravděpodobností 70% nebude fungovat ani zítra.

Stav sítě tedy nabývá dvou hodnot, A_1 =„funguje" a A_2 =„nefunguje". Pravděpodobnosti v čase k lze popsat dvourozměrným vektorem

$$\boldsymbol{\pi}^{(k)} = (\pi_1^{(k)}, \pi_2^{(k)}).$$

Matice přechodu je v tomto případě

$$\mathbf{W} = \begin{pmatrix} 0.6 & 0.4 \\ 0.3 & 0.7 \end{pmatrix}.$$

Jestliže tedy včera byl stav sítě popsán vektorem $\boldsymbol{\pi}^{(1)}$, pak pro dnešní stav platí:

$$\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{W}\,.$$

Tedy např. je-li $\boldsymbol{\pi}^{(1)} = (1, 0)$, pak $\boldsymbol{\pi}^{(2)} = (0.6, 0.4)$, a je-li $\boldsymbol{\pi}^{(1)} = (0, 1)$, pak $\boldsymbol{\pi}^{(2)} = (0.3, 0.7)$. Takto mohu pokračovat a druhý den mám rozložení

$$\boldsymbol{\pi}^{(3)} = \boldsymbol{\pi}^{(2)} \cdot \mathbf{W} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{W}^2$$

[1] Toto tvrzení není přesné, protože pravděpodobnost realizace *nekonečné* posloupnosti bude nejspíš nula. Čtenář nechť si představí konečnou vybranou posloupnost s obecně nenulovou pravděpodobností.

a tedy

$$\boldsymbol{\pi}^{(3)} = \begin{cases} (0.48, 0.52), & \text{je-li } \boldsymbol{\pi}^{(1)} = (1,0), \\ (0.39, 0.61), & \text{je-li } \boldsymbol{\pi}^{(1)} = (0,1). \end{cases}$$

Pokračuji-li dále, zjistím, že rozložení $\boldsymbol{\pi}^{(n)}$ pro velká n nebude už vůbec záviset na $\boldsymbol{\pi}^{(1)}$ a dostanu tzv. limitní rozložení

$$\lim_{n\to\infty} \boldsymbol{\pi}^{(n)} = \boldsymbol{\pi} \doteq (0.4286, 0.5714)\,, \tag{4.8}$$

tedy průměrná pravděpodobnost fungování je 43%. Opravdu je tomu tak ale vždycky? Zkusme něco jiného. Byl jsem na dovolené a nevím, zda včera síť fungovala nebo ne. Jeden kolega říká, že síť nefungovala, druhý tvrdí, že ano. Pravděpodobnosti pro a proti jsou tedy asi stejné a počáteční stav proto vezmeme ve tvaru

$$\boldsymbol{\pi}^{(1)} = (0.5, 0.5).$$

Jaká je tedy pravděpodobnost, že síť bude fungovat, až dnes přijdu do práce?

$$\boldsymbol{\pi}^{(2)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{W} = (0.45, 0.55)$$

a tedy pravděpodobnost je 45%. Budu-li nyní pokračovat ve výpočtu pravděpodobnosti dál, zjistím, že po několika dnech dostanu opět rozložení (4.8).

Ve statistické fyzice nás zajímá měření veličin. Zde může jako příklad veličiny (pozorovatelné) sloužit výdělek: jestliže síť funguje, vydělám $X(\text{„funguje“}) = 2000\,\text{Kč}$ za den, jestliže nefunguje, nemohu pracovat a beru pouze $X(\text{„nefunguje“}) = 500\,\text{Kč}$. Průměrný výdělek je dán střední hodnotou

$$\langle X \rangle = \pi(\text{„funguje“})X(\text{„funguje“}) + \pi(\text{„nefunguje“})X(\text{„nefunguje“}) \doteq 1143\,\text{Kč}\,.$$

Matice přechodu v našem jednoduchém příkladu má tu vlastnost, že systém ztrácí „paměť“, tj. po jisté době pravděpodobnost π_j nalezení jevu A_j vůbec nezávisí na tom, z jakého rozložení jsme vyšli. A jak tento příklad souvisí s původním problémem vybrané posloupnosti, s níž budeme procházet konfigurační prostor? Jednoduše. Mějme posloupnost stavů čili nyní konfigurací $\{A^{(k)}\}_{k=1}^{n}$ vybranou z Markovova řetězce $\{\mathcal{S}^{(1)}, \mathcal{S}^{(2)}, \ldots\}$ s limitním rozložením

$$\pi_j = \frac{\exp(-\beta U_j)}{\sum_{k=1}^{M} \exp(-\beta U_k)} \equiv \frac{\exp(-\beta U_j)}{Q}\,, \tag{4.9}$$

kde jsme označili $U_j = U(A_j)$. Pak střední hodnota veličiny X podél tohoto řetězce,

$$\bar{X} = \frac{1}{n} \sum_{k=1}^{n} X^{(k)}\,, \tag{4.10}$$

kde $X^{(k)} = X(A^{(k)})$, bude pro rostoucí n konvergovat k souborové střední hodnotě dané vztahem

$$\langle X \rangle = \sum_{j=1}^{M} \pi_j X(A_j) \equiv \sum_{j=1}^{M} \pi_j X_j\,. \tag{4.11}$$

Zbývá určit podmínky, za kterých $\boldsymbol{\pi}^{(k)} = \boldsymbol{\pi}^{(1)} \cdot \mathbf{W}^k$ konverguje k limitnímu rozložení $\boldsymbol{\pi}$. Jestliže:

1. všechny stavy jsou dosažitelné z libovolného stavu v konečném čase s nenulovou pravděpodobností a
2. žádný stav není periodický (stav A_i je periodický, jestliže existuje perioda m taková, že je-li $\pi_i^{(k)} = 0$, pak $\pi_i^{(k+m)} = 0$ a je-li $\pi_i^{(k)} \neq 0$, pak $\pi_i^{(k+m)} \neq 0$),

pak se množina stavů nazývá *ergodická* a pro libovolné počáteční rozložení pravděpodobností $\boldsymbol{\pi}^{(1)}$ existuje limita $\boldsymbol{\pi} = \lim_{k\to\infty} \boldsymbol{\pi}^{(k)}$ [33, 34]. Rozložení pravděpodobnosti $\boldsymbol{\pi}$ je tedy řešením rovnice

$$\boldsymbol{\pi} \cdot \mathbf{W} = \boldsymbol{\pi} \tag{4.12}$$

a toto řešení je jediné.

■ Jinými slovy, vektor stavů $\boldsymbol{\pi}$ je vlastním levým vektorem stochastické matice $\mathbf{W}$. Lze ukázat, že všechna další vlastní čísla jsou v absolutní hodnotě menší než 1.

4.2.2 Určení matice přechodu

Potřebujeme tedy zkonstruovat posloupnost (Markovův řetězec) konfigurací tak, aby se pravděpodobnost výskytu jednotlivých konfigurací rovnala Boltzmannově váze (4.9), která tak bude představovat limitní rozložení jisté, zatím neznámé matice přechodu. Toto je právě opačný problém než ten, který se obvykle řeší v teorii Markovových řetězců, tj. k dané matici přechodu nalézt limitní rozdělení.

Pro určení matice přechodu máme celkem tři podmínky:

$$W_{i\to j} \geq 0 \qquad \text{pro všechna } i,j = 1,\ldots,M \tag{4.13}$$

$$\sum_{j=1}^{M} W_{i\to j} = 1 \qquad \text{pro všechna } i = 1,\ldots,M \tag{4.14}$$

$$\boldsymbol{\pi} \cdot \mathbf{W} = \boldsymbol{\pi}\,. \tag{4.15}$$

Poslední rovnice vyjadřuje podmínku tzv. detailní rovnováhy. Toto je nutná podmínka zajišťující vlastnosti limitního (stacionárního) rozložení pravděpodobnosti. Umíme ji splnit, požadujeme-li silnější podmínku, tzv. podmínku mikroskopické reverzibility. Stačí, aby

$$\pi_i W_{i\to j} = \pi_j W_{j\to i} \tag{4.16}$$

a podmínka detailní rovnováhy je splněna.

■ Platnost podmínky (4.15) za předpokladu platnosti (4.16) snadno ukážeme, aplikujeme-li operátor $\sum_{i=1}^{M}$ na obě strany rovnice (4.16) a dosadíme-li jedničku za $\sum_{i=1}^{M} W_{j\to i}$ (viz rov. (4.14)) na pravé straně.

Rovnice (4.13)–(4.15) neurčují matici přechodu jednoznačně. Soustava (4.14) a (4.15) totiž představuje celkem $2M$ rovnic pro $M \times M$ neznámých.

Využijme nyní toho, že $\pi_i = \exp(-\beta U_i)/Q$ je limitní rozložení. Po dosazení do (4.16) dostaneme buď

$$\frac{W_{j\to i}}{W_{i\to j}} = \frac{\pi_i}{\pi_j} = \exp[-\beta(U_i - U_j)]\,, \tag{4.17}$$

nebo $W_{j\to i} = W_{i\to j} = 0$. Jak je vidět, podařilo se nám zbavit, alespoň v poměru pravděpodobností, neznámé funkce Q; v obecném případě (jsou-li π_i jiné než Boltzmannovy pravděpodobnosti) tento výsledek znamená, že nám stačí zadat jen relativní pravděpodobnosti a nemusíme se starat o součet čísel π_i. Snadno se nyní přesvědčíme, že matice přechodu definovaná vztahy

$$W_{i\to j} = \begin{cases} \alpha_{i\to j} & \text{pro } i \neq j \text{ a } \pi_j \geq \pi_i, \\ \alpha_{i\to j}\dfrac{\pi_j}{\pi_i} & \text{pro } i \neq j \text{ a } \pi_j < \pi_i, \\ 1 - \displaystyle\sum_{k,k\neq i} W_{i\to k} & \text{pro } i = j, \end{cases} \tag{4.18}$$

kde $\alpha_{i\to j}$ je libovolná *symetrická* stochastická (tj. platí pro ni (4.13) a (4.14)) matice, splňuje podmínku mikroskopické reverzibility. Toto je matice navržená Metropolisem (viz citace[1] na str. 12) a používaná dodnes. Poznamenejme ještě, že první dva řádky v (4.18) lze zapsat kompaktně jako

$$W_{i\to j} = \alpha_{i\to j} \min\left\{1, \frac{\pi_j}{\pi_i}\right\} \quad \text{pro } i \neq j\,. \tag{4.19}$$

■ Metoda (4.18) není jediná. Symetrická matice známá jako Barkerova[2] (někdy též Glauberova[3]) má tvar

$$W_{i\to j} = \begin{cases} \alpha_{i\to j}\dfrac{\pi_j}{\pi_i + \pi_j} & \text{pro } i \neq j\,, \\ 1 - \displaystyle\sum_{k,k\neq i} W_{ik} & \text{pro } i = j\,, \end{cases} \tag{4.20}$$

je však sotva kdy lepší než Metropolisova.

Jinou možností, používanou především pro mřížkové modely s krátkodosahovým potenciálem, je tzv. metoda tepelné lázně (heat-bath method[4]). Předpokládá se, že umíme spočítat částečnou statistickou sumu pro jistou podmnožinu $\mathcal{C}_{\text{part}}$ celého konfiguračního prostoru $\mathcal{C}$, který předpokládáme diskrétní. Matice přechodu metody tepelné lázně je

$$W_{i\to j} = \exp(-\beta U_j) \Big/ \sum_{A_k \in \mathcal{C}_{\text{part}}} \exp(-\beta U_k) \quad \text{pro } A_i, A_j \in \mathcal{C}_{\text{part}}\,. \tag{4.21}$$

Abychom byli konkrétní, zvolme si za $\mathcal{C}_{\text{part}}$ množinu všech konfigurací lišících se jen hodnotou proměnné na jednom mřížkovém bodu b. Tuto proměnnou (spin) označíme s_b. Pak lze napsat

$$W(s_b \to s_b') = \exp[-\beta U(s_b')] \Big/ \sum_{s_b^*} \exp[-\beta U(s_b^*)]\,, \tag{4.22}$$

kde závislost veličin na ostatních spinech, které se v daném kroku nemění, již nevyznačujeme. Konfigurace spinu s_b vybíráme tedy s Boltzmannovou váhou danou energiemi konfigurací a teplotou, což si lze představit jako ponoření spinu do termostatu či „tepelné lázně" o dané teplotě T. Spin tak „zapomene" na předchozí stav a $W_{i\to j}$ nezávisí na A_i. Nyní předběhneme poněkud výklad a srovnáme metodu tepelné lázně s Metropolisovou metodou. Metoda tepelné lázně je vhodnou alternativou, jestliže Metropolisova metoda se stane neefektivní pro příliš mnoho odmítnutých konfigurací. Implementace je však náročnější a zpravidla je založena na tom, že pro všechna možná okolí jednoho spinu předem vypočteme a tabelujeme rozložení (4.21). K výběru spinu pak použijeme např. algoritmus 12.2, viz odd. 12.1.2. Pro systém s pouze dvěma hodnotami spinu (Isingův model, mřížkový plyn) je metoda tepelné lázně totožná s Barkerovou metodou.

---

[2] A. A. Barker: *Aust. J. Phys.* **18**, 119, (1965).
[3] R. J. Glauber: *J. Math. Phys* **4**, 294 (1963).
[4] M. Creutz, L. Jacobs a C. Rebbi: *Phys. Rev. Lett.* **42**, 1390 (1979).

4.2.3 Zkušební změna konfigurace

Zbývá odpovědět na otázku, co je matice $\alpha_{i\to j}$ vystupující v maticích přechodu (4.18), případně (4.20). Tato matice udává podmíněnou pravděpodobnost generování *zkušební* konfigurace A_j z konfigurace A_i; v následujícím oddíle se dozvíme, kdy bude tato zkušební konfigurace přijata a kdy odmítnuta. Pro klasický spojitý systém budeme mít místo matice funkci $\alpha(\mathbf{r}^N \to \mathbf{r}'^N)$.

Ve standardní Metropolisově metodě je tato matice symetrická. Máme-li tedy konfiguraci A_i, pak konfiguraci A_j musíme vygenerovat s pravděpodobností stejnou, jakou bychom z konfigurace A_j generovali konfiguraci A_i. Podobně je tomu ve spojitém případě, kde pravděpodobnost je nahrazena hustotou pravděpodobnosti; nesmíme jen zapomenout na to, že tato hustota pravděpodobnosti je definovaná vzhledem ke kartézským proměnným $\mathbf{r}^N$ a může se změnit, jestliže bychom chtěli přejít k jiným souřadnicím (např. sférickým). Dále musíme mít při návrhu zkušebního posunutí na paměti ergodičnost, tedy aby systém mohl (pomocí co nejmenšího počtu kroků) přecházet z jedné části konfiguračního prostoru do jiné. Speciálně ve spojitých modelech je nutno dát pozor na to, aby systém snadno překonal velké potenciálové bariéry, např. u vnitřních stupňů volnosti. Podobně u některých mřížkových modelů může nastat situace, kdy není možné jednu konfiguraci přeměnit na druhou pouze postupnými záměnami jednotlivých spinů; pak je nutné měnit v jednom kroku celou skupinu spinů najednou.

Pro konkrétnost uvedeme dva typické příklady.

Mřížkový systém s diskrétním konfiguračním prostorem

Uvažujme pro jednoduchost systém, jehož proměnné na každém vrcholu mřížky nabývají pouze tří hodnot, které označíme $\{1, 2, 3\}$. V jednom kroku MC simulace budeme měnit hodnotu pouze na jednom vrcholu mřížky[5]. Máme-li na něm hodnotu 1, zkusíme ji změnit s pravděpodobností 1/2 na 2 a s touž pravděpodobností na 3, podobně 2 změníme s pravděpodobností 1/2 na 1 či 3 a 3 na 1 či 2.

■ Matice α pro výše uvedený příklad je obrovská – $3^N \times 3^N$, kde N je počet vrcholů mřížky. Pokusme se ji zapsat alespoň pro $N = 2$, tedy pro mřížtičku $\boxed{s_1}\boxed{s_2}$, a pro zkušební změnu spinu s_2:

	1 1	1 2	1 3	2 1	2 2	2 3	3 1	3 2	3 3
1 1	0	1/2	1/2	0	0	0	0	0	0
1 2	1/2	0	1/2	0	0	0	0	0	0
1 3	1/2	1/2	0	0	0	0	0	0	0
2 1	0	0	0	0	1/2	1/2	0	0	0
2 2	0	0	0	1/2	0	1/2	0	0	0
2 3	0	0	0	1/2	1/2	0	0	0	0
3 1	0	0	0	0	0	0	0	1/2	1/2
3 2	0	0	0	0	0	0	1/2	0	1/2
3 3	0	0	0	0	0	0	1/2	1/2	0

(4.23)

Pro napsání počítačového programu naštěstí nepotřebujeme zapsat explicitně ani matici α ani $\mathbf{W}$ a stačí nám implicitní vyjádření pomocí algoritmu.

[5] Je možné a někdy i výhodné měnit více spinů najednou, např. ve čtverci 2×2 – podmínkou je, aby pravděpodobnosti přijetí v Metropolisově testu byly dost velké.

Spojitý klasický systém

V případě spojitého systému se podmínka symetrie $\alpha(\mathbf{r}^N \to \mathbf{r}'^N) = \alpha(\mathbf{r}'^N \to \mathbf{r}^N)$ nezmění, pokud zůstaneme u kartézských souřadnic částic. Ve většině simulací se v jednom kroku hýbe pouze s jednou částicí, protože, jak víme z předběžných úvah, se nová konfigurace nesmí příliš lišit od původní, jinak bychom opět dostali velmi nepravděpodobnou konfiguraci (např. překryv molekul), která by nemohla být přijata. Nejjednodušší metodou je přičtení náhodného čísla $u_{(-d,d)}$ rovnoměrně rozloženého v intervalu $(-d, d)$ ke každé kartézské souřadnici vybrané částice o poloze $\mathbf{r} = (r_x, r_y, r_z)$.

$$\begin{aligned} r'_x &= r_x + u_{(-d,d)} , \\ r'_y &= r_y + u_{(-d,d)} , \\ r'_z &= r_z + u_{(-d,d)} . \end{aligned} \tag{4.24}$$

4.2.4 Realizace kroku Metropolisovy metody

Vzorec (4.19) znamená, že změna zkušební konfigurace daná prvkem matice $\alpha_{i\to j}$ bude přijata s pravděpodobností

$$p_{\text{přij}} = \min\left\{1, \frac{\pi_j}{\pi_i}\right\} , \tag{4.25}$$

což v případě kanonického rozložení značí

$$p_{\text{přij}} = \min\left\{1, \exp(-\beta\Delta U)\right\} , \tag{4.26}$$

kde

$$\Delta U = U(A_j) - U(A_i) \equiv U_j - U_i . \tag{4.27}$$

Jinými slovy, máme-li v kroku k konfiguraci A_i, tj. $\mathcal{S}^{(k)} = A_i$, pak budeme mít v kroku $k+1$ někdy novou konfiguraci A_j a někdy starou A_i: abychom zestručnili zápis algoritmů (s vědomím toho, že explicitní zápis matic $\alpha_{i\to j}$ a $W_{i\to j}$ nepotřebujeme), budeme značit výchozí konfiguraci $A^{(k)}$, zkušební konfiguraci A^{zkus}, a výslednou konfiguraci po provedení kroku $A^{(k+1)}$.

Nyní tedy již můžeme shrnout schéma jednoho Monte Carlo kroku pro nejběžnější simulaci Metropolisovou metodou:

1. Zvolíme částici, kterou se bude hýbat, mřížkový bod a podobně. Částice je možné procházet systematicky v cyklu (jsou-li stejného druhu) nebo vybírat náhodně.
2. Vypočteme zkušební konfiguraci A^{zkus} tak, že změníme náhodně polohu (orientaci, spin) vybrané částice, viz odd. 4.2.3.
3. Vypočteme $\Delta U = U(A^{\text{zkus}}) - U(A^{(k)}) \equiv U^{\text{zkus}} - U^{(k)}$.
4. Konfiguraci přijmeme ($A^{(k+1)} = A^{\text{zkus}}$) s pravděpodobností $p_{\text{přij}}$ podle (4.26), zatímco s opačnou pravděpodobností $1 - p_{\text{přij}}$ ji odmítneme ($A^{(k+1)} = A^{(k)}$). To v řeči počítačového programu znamená, že vygenerujeme náhodné číslo $u = u_{(0,1)}$, viz rov. (12.4). Je-li $u < p_{\text{přij}}$, pak zkušební konfiguraci přijmeme, v opačném případě pokračujeme se starou konfigurací.
5. Pokračujeme bodem 1.

A to vše děláme tak dlouho, až získáme dostatečný počet konfigurací potřebných k výpočtu fyzikálních veličin.

■ Bod 4. má několik variant. Za prvé nemusíme počítat minimum (4.26) (nebo (4.25)), stačí když

4. Generujeme $u = u_{(0,1)}$. Je-li $u < \exp(-\beta\Delta U)$, pak zkušební konfiguraci přijmeme, v opačném případě pokračujeme se starou konfigurací.

Za druhé, chceme-li ušetřit generování náhodného čísla, všimneme si, že pro $\Delta U < 0$ (zkušební konfigurace je energeticky výhodnější) je $p_{\text{přij}} = 1$, a tedy zkušební konfigurace je vždy přijata. Podmínka nyní zní

4. Je-li $\Delta U < 0$, pak zkušební konfiguraci přijmeme a pokračujeme bodem 1., v opačném případě generujeme $u = u_{(0,1)}$ a je-li $u < \exp(-\beta\Delta U)$, pak zkušební konfiguraci přijmeme, v opačném případě pokračujeme se starou konfigurací.

Příklady implementace Metropolisovy metody jsou uvedeny v algoritmech 4.2 a 4.3.

4.3 Zlomek přijetí a nastavení parametrů

Důležitou charakteristikou MC simulací je tzv. zlomek přijetí (acceptance ratio), čili poměr

$$\chi = \frac{\text{počet přijatých konfigurací}}{\text{počet všech generovaných konfigurací}}. \tag{4.28}$$

Uvažujeme-li simulace klasických spojitých systémů, bude tento poměr zřejmě záviset na velikosti zkušebního posunutí d v rovnici (4.24) či v jiné podobné. Bude-li např. d příliš malé, bude změna energie v jednom kroku malá, Boltzmannův faktor $\exp(-\beta\Delta U)$ blízký jedničce, téměř každá konfigurace bude přijata a χ bude blízké jedničce, ale také efektivita simulace bude nízká, protože konfigurace se od sebe málo liší. Bude-li d velké, bude změna energie v jednom kroku také velká a většinou kladná, Boltzmannův faktor $\exp(-\beta\Delta U)$ blízký nule a téměř každá konfigurace bude odmítnuta, což je zřejmě také neefektivní. Existuje proto jistá optimální hodnota zlomku přijatých konfigurací, o které se dlouho naivně předpokládalo, že musí být blízká 1/2, novější výzkumy [15] však ukazují, že toto optimum leží ve většině případů okolo $\chi = 1/3$.

■ Přesněji řečeno, je nutno optimalizovat nejen délku kroku podle (4.24), ale celý tvar (pravděpodobnostní rozložení) zkušebního posunutí[6]. Již jen fyzikální úvaha vede k poznatku, že rovnoměrné rozložení v krychli podle (4.24) asi nebude pro homogenní a izotropní tekutiny tím nejlepším a že zkušební posunutí by mělo být sféricky symetrické. Nadto je neefektivní mít příliš krátká posunutí zároveň s dlouhými, jak je tomu při použití (4.24) (s jistou pravděpodobností je $\mathbf{r}'$ velmi blízké $\mathbf{r}$); tato úvaha však nemusí platit pro složité systémy, kde naopak můžeme potřebovat různé délky pro vzorkování různých stupňů volnosti. Lze proto doporučit např. rovnoměrné rozložení na povrchu koule či v kouli s vyjmutou koulí o menším poloměru.

Je možné také uvažovat nesymetrickou matici $\alpha_{i\to j}$, o tom však blíže v odd. 8.1.

Ve speciálních případech může být optimální χ jiné. Např. pro tekutiny tuhých těles je mnohem nižší (až 0.1), což ovšem souvisí s typickým způsobem implementace. Jiným příkladem jsou řídké systémy, zvláště blízko kritického bodu: zde je efektivnější mít délku zkušebního posunutí srovnatelnou s velikostí simulovaného systému i za cenu velmi malého zlomku přijatých konfigurací (třeba 1%) než malou změnu konfigurace v jednom kroku.

[6] J. Kolafa: *Mol. Phys.* **63**, 559 (1988).

```
INTEGER N po et atom
VECTOR r[1..N] konfigurace {r_i}_{i=1}^N
REAL Epair[1..N,1..N] tabulka párových energií
REAL Ezkus[1..N] párové energie zkušební konfigurace
REAL Utot, Uzkus, U
REAL u(VECTOR r1,r2) funkce vracející párovou energii u(|r2 - r1|) atomů s polohami r1 a r2
                     (vzhledem k okrajovým podmínkám - viz algoritmus 5.1 nebo 5.2)
REAL rnd() funkce vracející náhodné číslo v intervalu (0,1)
INTEGER i, j
REAL d velikost zkušebního posunutí

Předem: naplnění tabulky párových energií
FOR i := 1 TO N-1 DO
   FOR j := i+1 TO N DO nyní (i,j) prochází všechny páry částic, i < j
      Epair[i,j] := u(r[i],r[j])
      Epair[j,i] := Epair[i,j]

Jeden cyklus MC simulace (1 MC krok s každým atomem)
Utot := 0
FOR i := 1 TO N DO
   zkušební posunutí atomu i
   VECTOR rzkus
   rzkus.x := r[i].x + d*(rnd()-0.5)
   rzkus.y := r[i].y + d*(rnd()-0.5)
   rzkus.z := r[i].z + d*(rnd()-0.5)

   výpočet staré a nové energie: součet přes všechny atomy různé od i
   Uzkus := 0
   U := 0
   FOR j := 1 TO N DO IF j <> i THEN
      Ezkus[j] := u(r[j],rzkus)
      Uzkus := Uzkus + Ezkus[j]
      U := U + Epair[i,j]

   Metropolisův test; k je Boltzmannova konstanta
   IF rnd() < exp(-(Uzkus-U)/(k*T)) THEN
      konfigurace přijata
      r[i] := rzkus
      Utot := Utot + Uzkus
      FOR j := 1 to N do
         Epair[j,i] := Ezkus[j]
         Epair[i,j] := Ezkus[j]
   ELSE konfigurace odmítnuta
      Utot := Utot + U

PRINT "energie konfigurace =", Utot/2
příspěvek od každého páru jsme započetli dvakrát, proto Utot/2
```

Algoritmus 4.2: Metropolisova MC simulace N atomů a výpočet vnitřní energie.

```
INTEGER L velikost hrany mřížky
SPINTYPE s[1..L,1..L] konfigurace s_ij; SPINTYPE je typ spinové proměnné
REAL u(SPINTYPE s1,s2) funkce vracející interakční energii nejbližších sousedů
INTEGER i, j
INTEGER plus[1..L], minus[1..L] indexy sousedů
REAL Uzkus, U

Předem: indexy sousedů
FOR i := 1 TO L-1 DO plus[i] := i+1
plus[L] := 1
FOR i := 2 TO L DO minus[i] := i-1
minus[1] := L

jeden cyklus MC simulace (1 MC krok s každým spinem)
FOR i := 1 TO L DO
   FOR j := 1 TO L DO
      zkušební změna spinu na vrcholu i
      SPINTYPE szkus
      szkus := {nová náhodná hodnota}
         jiná možnost: szkus := s[i,j] + {malá symetrická změna}
         pro dvoustavový Isingův feromagnet stačí: szkus := -s[i,j]

      výpočet staré a nové energie: součet přes všechny nejbližší sousedy
      U := u(s[i,j],s[plus[i],j]) + u(s[i,j],s[minus[i],j])
          + u(s[i,j],s[i,plus[j]]) + u(s[i,j],s[i,minus[j]])
      Uzkus := u(szkus,s[plus[i],j]) + u(szkus,s[minus[i],j])
          + u(szkus,s[i,plus[j]]) + u(szkus,s[i,minus[j]]).

      Metropolisův test; k je Boltzmannova konstanta
      IF rnd() < exp(-(Uzkus-U)/(k*T)) THEN
         konfigurace přijata
         s[i,j] := szkus
```

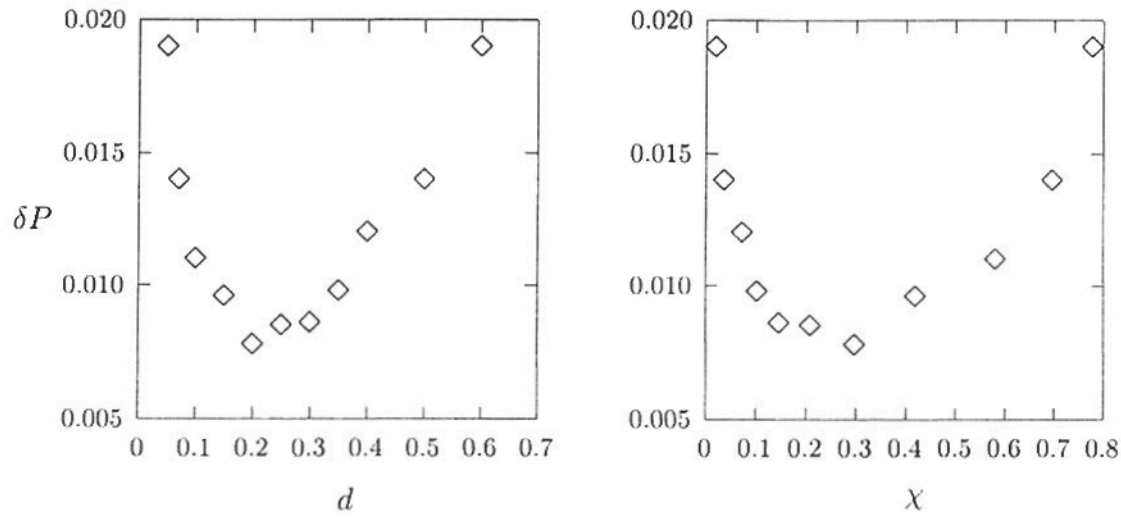
Algoritmus 4.3: Metropolisova MC simulace na periodické čtvercové mřížce s interakcí nejbližších sousedů.

Obrázek 4.1 ukazuje závislost chyby měření tlaku (podle viriálové stavové rovnice (5.12)) pro systém 64 Lennard-Jonesových atomů v periodických okrajových podmínkách na délce zkušebního posunutí a zlomku přijatých konfigurací. Vidíme, že optimální χ leží v intervalu $[0.2, 0.3]$.

■ Symetrickou matici zkušebních posunutí je možné získat i pomocí nějaké časově přesně reverzibilní metody MD, např. Verletovy. Protože bude stejně proveden Metropolisův test, nezáleží na kvalitě integrace a je možné použít dlouhé integrační kroky. Tato metoda, v které se mění najednou polohy všech částic, se nazývá *hybridní Monte Carlo*[7] a je pro nepříliš velké systémy překvapivě efektivní.

U mřížkových systémů s diskrétními stupni volnosti zpravidla nemáme simulační parametr, jímž bychom mohli zlomek přijetí ovlivnit. Ten však závisí na fyzikálních podmínkách. Např. při nízkých teplotách bude Boltzmannův faktor $\exp(-\beta\Delta U)$ velmi malý a simulace bude neefektivní. Někdy lze tento problém zmírnit např. výše zmíněnou metodou

[7] S. Duane, A. D. Kennedy, B. J. Pendleton a D. Roweth: *Physics Letters B* **195**, 216 (1987).



Obrázek 4.1: Závislost směrodatné chyby měření viriálového tlaku δP na maximální délce zkušebního posunutí d (vlevo) a na zlomku přijatých konfigurací χ (vpravo). Každý bod byl získán z 10^7 MC kroků (krok = pokus pohnout jednou částicí) na systému 64 LJ atomů ($\sigma = 1$, $\epsilon = 1$) při $T = 1.2$, $\rho = 0.8$.

tepelné lázně, zvláště jestliže měníme v jednom kroku celou skupinu spinů. Jiný algoritmus spočívá ve vytvoření seznamu všech možných elementárních nebo jednokrokových změn konfigurace. Z nich se pak vybere jedna na základě srovnání jejich pravděpodobností. Vybraná změna je pak vždy přijata. Varianta tohoto algoritmu bez neúspěšných pokusů bude popsána v kap. 9.2.

Pro speciální mřížkové systémy existují i tzv. klastrové algoritmy[8], kde se mění celá nepravidelná oblast spinů, tyto postupy jdou však již mimo rámec Metropolisovy metody a těchto skript.

[8] Viz [15] nebo R. H. Swendsen, J.-S. Wang a A. M. Ferrenberg v [18].